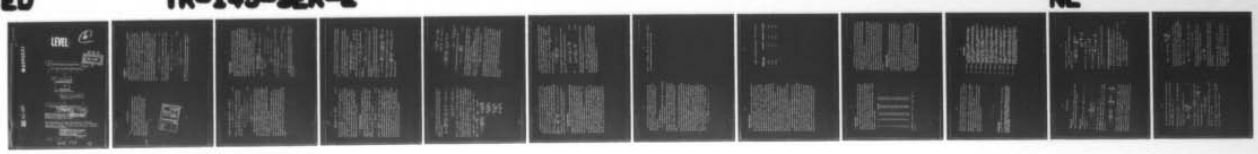
AD-A069 531 PRINCETON UNIV N J DEPT OF STATISTICS F/G 12/1
SIMULTANEOUS POISSON ESTIMATORS FOR A PRIORI HYPOTHESES ABOUT M--ETC(U)
MAR 79 H M HUDSON, K TSUI N00014-75-C-0453
UNCLASSIFIED TR-145-SER-2 NL

| OF |

AD
A069531



END
DATE
FILMED
7 -79
DDC

LEVEL II

(6)

6

SIMULTANEOUS POISSON ESTIMATORS
FOR A *PRIORI* HYPOTHESES ABOUT MEANS

by

10

H. Malcolm Hudson*
[redacted]ity

and

Kam-Wah Tsui†
Simon Fraser University

14 TR-145-SER-2

9

Technical Report No. 145, Series 2
Department of Statistics
Princeton University
11 March 1979

12 13p.

15

Research sponsored in part by a contract with the Office of Naval Research, No. N00014-75-C-0453, awarded to the Department of Statistics, Princeton University, Princeton, New Jersey.

- - - - - - - - - - - - - - - -

*Visiting Lecturer, on leave from the School of Economic and Financial Studies, Macquarie University, N.S.W., Australia.

†Department of Mathematics, Simon Fraser University, Vancouver, British Columbia, Canada.

A B S T R A C T

Estimators are presented for the simultaneous estimation of p means of Poisson variables with the properties of significant reduction in mean square error of estimation in the region of a pre-specified set of values and, if $p \geq 3$, safety, in the sense that mean square error is never worse than that of the minimum variance unbiased estimator.

7 or =

1. INTRODUCTION

Suppose $x_1, \ldots, x_p$ are p observations from p independent Poisson distributions with means $\lambda_1, \ldots, \lambda_p$, respectively. In estimating $\lambda = (\lambda_1, \ldots, \lambda_p)$, Clevenson and Zidek (1975), Leonard (1976), and Peng (1978) have shown in different ways that the usual estimator $x = (x_1, \ldots, x_p)$ is inadequate. In particular, Peng's result shows that x is inadmissible under squared error loss $L(\lambda, \hat{\lambda}) = \sum_{i=1}^{p} (\lambda_i - \hat{\lambda}_i)^2$, when $p \geq 3$. However, his proposed estimators perform well only when the λ_i's are small. K. Tsui (1978) extends Peng's result and provides a large class of estimators of the Poisson means which can incorporate some knowledge of parameter values in order to obtain further improvement over the usual estimator x. For each nonnegative integer k, one of Tsui's estimators is of the form

$$\hat{\lambda}_i = x_i - r_k h(x_i)/S, \quad i = 1, \ldots, p, \tag{1.1}$$

where $S = \sum_{j=1}^{p} h^2(x_j)$; r_k is the maximum of (i) the number of x_i greater than k, less two and (ii) zero; and, for some $b \geq 0$, $h(y)$ is defined on the positive integers as:

$$\begin{aligned} h(y) &= -b, \quad \text{if } k > 0 \text{ and } y = 0, 1, \ldots, k-1 \\ &= 0, \quad \text{if } y = k \\ &= 1, \quad \text{if } y = k+1 \\ &= 1 + \sum_{n=k+2}^{y} n^{-1}, \quad \text{if } y = k+2, k+3, \ldots \end{aligned} \tag{1.2}$$

The estimator (1.1) is designed to provide maximum advantage when the λ_i's are close to the predetermined integer k. Adaptive

estimators are also proposed by Tsui. An example is

$$\hat{\lambda}_i = x_i - r_m h_i(x)/S' , \quad i = 1, \ldots, p , \tag{1.3}$$

where $S' = \sum_{i=1}^{p} h_i^2(x)$; r_m is the maximum of (i) the number of x_j greater than the minimum of the x_j, less two and (ii) zero; and $h_i(x)$, $i = 1, \ldots, p$, are defined as

$$\begin{aligned} h_i(x) &= 0 , \quad \text{if } x_i = \min(x_j) = m \text{ or } x_i < 0 \\ &= 1 , \quad \text{if } x_i = m+1 \text{ and } m \geq 0 \\ &= 1 + \sum_{n=m+2}^{x_i} n^{-1} , \quad \text{if } x_i > m+1 \text{ and } m \geq 0 . \end{aligned} \tag{1.4}$$

These adaptive estimators possess better risk properties over a large region of the parameter space and are especially good when the λ_i's are close to one another. It is required that p be greater than 3 when these adaptive estimators are used. The improvements in risk for both estimators given in (1.1) and (1.3) exceed $E_\lambda(r_k^2/S)$ and $E_\lambda(r_m^2/S')$, respectively. More improvement can then be achieved by reducing S or S' if possible.

In this paper we propose a number of new estimators of the form (1.1) and (1.3), with modifications of the choices $h(y)$ and $h_i(y)$. Considerable improvements in risk properties are thus obtained, as demonstrated by the simulation results of Section 3 and the example of Section 4. Section 2 introduces these modifications, concluding with a discussion of variations of (1.1) that may be found useful.

2. THE MAIN RESULTS

In this section, we propose three estimators of the Poisson means which are modified versions of (1.1) and (1.3). The first estimator pulls the usual estimator towards a prefixed nonnegative integer k, when p is at least four. The second estimator shrinks the usual estimator towards a point determined by the data, namely, the minimum of the observations; p is required to be at least five in this case. Under squared error loss, these estimators have risk uniformly smaller than that of the usual estimator.

Let N_j = the number of observations equal to j, and ℓ denote the largest observation. Set $r_k^* = \max(0, (\sum_{j=k+1}^{\ell} N_j - 3))$.

2.1 Pulling x towards k

For a fixed nonnegative integer k and some $b \geq 0$, define $h^*(y)$ as follows:

$$\begin{aligned} h^*(y) &= -b , \quad \text{if } y = 0, 1, \ldots, k-1 \\ &= 0 , \quad \text{if } y = k \text{ or } y < 0 \\ &= \sum_{n=k+1}^{y} n^{-1} , \quad \text{if } y = k+1, k+2, \ldots . \end{aligned} \tag{2.1.1}$$

Let $S^* = \sum_{j=1}^{p} h^{*2}(x_j)$. Replacing h, r_k and S in (1.1) with h^*, r_k^* and S^*, we have a new estimator:

$$\begin{aligned} \hat{\lambda}_i^* &= x_i - r_k^* h^*(x_i)/S^* , \quad i = 1, \ldots, p, \text{ if } S^* \neq 0 , \\ \hat{\lambda}_i^* &= x_i , \quad i = 1, \ldots, p, \text{ if } S^* = 0 . \end{aligned} \tag{2.1.2}$$

Using squared error loss and following the argument of Lemma (5.6) of Hudson (1978), we have the risk improvement $R(\lambda,x) - R(\lambda,\hat{\lambda}^*)$ greater than or equal to

$$E_\lambda\left[kN_k r_k^* b/(S^* + b^2) + r_k^{*2}/S^*\right] \quad . \qquad (2.1.3)$$

When the Poisson means λ_i are large but close to one another, S^* is likely to be small if $\lambda_i \approx k$, $i = 1, \ldots, p$. In this case, S^* is much smaller than S as given in (1.1). Hence, the risk improvement using (2.1.2) is expected to be greater than that using (1.1) if $\lambda_i \approx k$ and k is not small. Appropriate choice of k, guided by the prior information available might lead to a substantial improvement in risk over the usual estimator. The nonnegative integer k should be chosen to be close to the prior mean.

2.2 <u>An adaptive estimator</u>

In most practical situations, an estimator which pulls x towards a point determined by the observations themselves and retains good risk properties is desirable. Define functions $H_i(x)$ to be $\sum_{n=m+1}^{x_i} n^{-1}$ if $x_i > m \equiv \min(x_j) \geq 0$, zero otherwise. Then, define the adaptive estimator with i^{th} coordinate

$$\hat{\lambda}_i = x_i - r_m' H_i(x)/S_*' \ , \quad i = 1, \ldots p \ , \qquad (2.2.1)$$

where $S_*' = \sum_{j=1}^{p} H_j^2(x)$, $r_m' = \max(0, p - N_m - 3)$, and N_m is the number x_j that are equal to the minimum m. This estimator shrinks x towards the smallest observation. The improvement in risk of (2.2.1) over the usual estimator can be shown to be at least $E_\lambda((r_m^*)^2/S)$ with the proof again along the line of Hudson (1978), lemma (5.6). S_*' will always be smaller than S' (as given in (1.3)) and is likely to be much smaller when k is at all large. Hence, the choice (2.2.1) will dominate the choice (1.3). The simulation results reported in Section 3 illustrate this performance.

Notice that (1.3) and (2.2.1) require different lower bounds on p, but this makes little difference if p is large, and see Section 2.4.

More adaptive estimators can be obtained by similar modifications of those suggested by Tsui, but we will not carry through the details here. One further result is of interest, however.

2.3 <u>Another estimator pulling x toward k</u>

Define $H(y) = \sum_{n=1}^{y} n^{-1}$, $y = 1, 2, \ldots$, and $H(0) = 0$. Taking a small liberty with notation, define $H_j = H(j) - H(k)$, for $j = 0, 1, 2, \ldots$. Set $S_H = \sum_{i=1}^{p} H_{x_i}^2 = \sum_{j=0}^{\ell} N_j H_j^2$, and define the estimator by

$$\hat{\lambda}_i = x_i - \frac{r_k^*}{S_H} H_{x_i} \ , \quad \text{for } i = 1, 2, \ldots, p \ . \qquad (2.3.1)$$

This estimator dominates x, while shrinking observations below k differentially.

To prove this result, we note that it is easy to show, following Stein's method of integration by parts, that the difference in risks of x and $\hat{\lambda}$ is

$$R(\lambda, x) - R(\lambda, \hat{\lambda}) = E\left\{\frac{2r_k^*}{S} \sum_{j=1}^{\ell} N_j \frac{S - H_j(H_j + H_{j-1})}{S - \frac{1}{j}(H_j + H_{j-1})} - \frac{(r_k^*)^2}{S}\right\}, \qquad (2.3.2)$$

where, for convenience, we have dropped the subscript H. We shall show that $\sum_{j=1}^{\ell} N_j \frac{S - H_j(H_j + H_{j-1})}{S - \frac{1}{j}(H_j + H_{j-1})}$ exceeds $\left(\sum_{j=k+1}^{\ell} N_j - 3\right)$, thus establishing the result.

In considering the summation of terms involving observations exceeding k, the argument used in Hudson (1978) continues to apply, and we may obtain

$$\sum_{j=k+1}^{\ell} N_j \frac{S-N_j(N_j+N_{j-1})}{S-\frac{1}{j}(N_j+N_{j-1})} \ge \left(\sum_{j=k+1}^{\ell} N_j\right) - 2\frac{S_+}{S} \quad \text{if} \quad N_\ell(N_\ell+N_{\ell-1}) < S$$

$$\ge \left(\sum_{j=k+1}^{\ell} N_j\right) - 3\frac{S_+}{S} \quad \text{if} \quad N_\ell(N_\ell+N_{\ell-1}) > S$$

where $S_+ = \sum_{j=k+1}^{\ell} N_j N_j^2$. For $j \le k$ we note that $S - \frac{1}{j}(N_j + N_{j-1}) > 0$, and $S - N_j(N_j + N_{j-1}) > 0$ except possibly for the smallest nonzero observation, which we shall denote by $j = m$, assumed less than $k + 1$. If in fact $S - N_m(N_m + N_{m-1}) > 0$, or $m > k$, the result is immediate on replacing terms of the summation below k by 0 and using the result above, noting $S_+/S \le 1$. If $N_m(N_m + N_{m-1}) > S$, then since $S_- \ge N_m N_m^2$ (where $S_- = \sum_{j=m}^{k} N_j N_j^2$), we have $(1+\beta)N_m^2 > S$ and $N_m N_m^2 < S_-$ where $\beta = \frac{N_{m-1}}{N_m} = 1 - \frac{1}{mN_m} \ge 1$. Therefore

$$0 > N_m \frac{S - N_m(N_m + N_{m-1})}{S - \frac{1}{m}(N_m + N_{m-1})} = N_m \frac{S - (1+\beta)N_m^2}{S - \frac{1}{m}N_m(1+\beta)}$$

$$\ge \frac{N_m S - (1+\beta)S_-}{S - \frac{1}{mN_m} N_m^2(1+\beta)}$$

$$\ge \frac{N_m S - (1+\beta)S_-}{S - \frac{1}{mN_m} S}$$

$$= \frac{N_m S - (1+\beta)S_-}{\beta S}$$

$$\ge \frac{S - (1+\beta)S_-}{\beta S} = \frac{1}{\beta} - \left(1 + \frac{1}{\beta}\right)\frac{S_-}{S}$$

$$= \frac{1}{\beta}\frac{S_+}{S} - \frac{S_-'}{S} \ge -\frac{S_-}{S} \quad \text{since } \beta \ge 1 .$$

$$\therefore \sum N_j \frac{S - N_j(N_j + N_{j-1})}{S - \frac{1}{j}(N_j + N_{j-1})} \ge \left(\sum_{j=k+1}^{\ell} N_j\right) - \frac{3S_+}{S} - \frac{S_-}{S}$$

$$\ge \sum_{j=k+1}^{\ell} N_j - 1 - \frac{2S_+}{S} \ge \left(\sum_{j=k+1}^{\ell} N_j - 3\right).$$

We thus have shown that the estimator (2.3.1) has risk

$$R(\mu,\hat{\lambda}) \le \Sigma\mu_i - E\left\{\frac{(r_k^*)^2}{S_N}\right\}. \tag{2.3.3}$$

This estimator is a discrete analogue of the Stein estimators developed for continuous exponential families by Hudson (1978). In continuous exponential families the identity

$$E((X-\mu)g(X)) = E(a(X)g'(X)) ,$$

where $a(\cdot)$ is a function dependent on the exponential family considered, is used to obtain risk properties for estimators of the form

$$x_i - \frac{(p-2)}{S_H} H(x_i) \quad \text{with} \quad H(x) = \int_k^x [a(x)]^{-1} dx .$$

The corresponding Poisson identity $E((X-\lambda)g(X)) = E(X[g(X) - g(X-1)])$ (see Hudson (1978)) may be used to motivate the choice (2.3.1), but not the choice of estimator given by (2.1.1) and (2.2.1). The choice r_k^* in (2.3.1) is smaller than one might hope; the authors expected r_0^* to be a more suitable choice. Nevertheless, we have little insight as to whether improvements increasing r_k^* are advantageous, except when k is large. See Section 2.4 for further discussion of this point.

In comparing the risk (2.3.3) with the risk $E((r_k^*)^2/S)$, given by the choice $b = 0$ in (2.1.3), it appears that since $S < S_H$ the estimator (2.1.1), with $b = 0$, should be preferred to that above. That this is not the case appears by comparison of the exact expression (2.3.2) with the inequality (2.3.3) in favorable circumstances where $X_1, \ldots, X_p$ are not distinguishable from independent Poisson observations with mean k. The exact expression will then approximate $\left[\frac{2r_k^*}{S_H}(2r_k^*) - \frac{(r_k^*)^2}{S_H}\right] = \frac{3(r_k^*)^2}{S_H}$ for large p, and S_H will be about twice the size of S. The estimator (2.3.1) might therefore be expected to be up to 50% better in risk savings than (2.1.1).

2.4 Concluding remarks

We mention in passing that by considerably more complicated arguments the choices r_k^* and r_m^* in Sections 2.1 - 2.3 can be replaced by r_k and r_m as defined in Section 1, provided the plus rule modification of the estimator, detailed below, is made. This allows one to apply the estimation procedures of Section 2 in precisely the same circumstances as the estimators of Section 1. The gain is slight for moderate or large p.

"Plus rule" versions of the above estimator, which will not shrink observations beyond k, can be obtained by redefining S as the greater of S, as defined above, and r_k^*/k. Alternatively, one may wish to enforce the restriction that S exceed the expected value of S when all parameters equal k, for S is a measure of the discrepancy of the observations from the hypothesized mean k. For large k this expected value is close to $p/2k$ for the estimator (2.1.1) with $b = 0$, and to p/k for the estimator (2.3.1).

We have discussed, in Section 2.3, the possibility of replacing the "constant" r_k^* by r_o^*, an adjustment that could be very beneficial if the resulting estimator still dominates n and substantially improves the risk near the prior guess k. To dominate x it is sufficient for $\sum_{j=m}^{\ell} N_j A_j$ to exceed $\frac{1}{2}\left(\sum_{j=m}^{\ell} N_j - 3\right)$ when this second term is positive; where

$$A_j = \frac{S - H_j(H_j + H_{j-1})}{S - \frac{1}{j}(H_j + H_{j-1})}, \quad j = 1, 2, \ldots, 1 \quad .$$

This result follows from appropriate revision of (2.3.2). When $\sum_{j=k+1}^{\ell} N_j \geq \frac{3S_+}{S}$ the condition above will be satisfied if

$$\sum_{j=k+1}^{\ell} N_j A_j \geq \left(\sum_{j=k+1}^{\ell} N_j - \frac{3S_+}{S}\right) \qquad (2.4.1)$$

and

$$\sum_{j=m}^{k} N_j A_j \geq \frac{1}{2}\left(\sum_{j=m}^{k} N_j - \frac{3S_-}{S}\right) \quad . \qquad (2.4.2)$$

A *proof of these results* is contained in the appendix, for an estimator obtained as a modification of (2.3.1).

The restriction $\sum_{j=k+1}^{\ell} N_j \geq 3S_+/S$ is negated only when the data is such that at most two observations exceed k, and these make a disproportionately large contribution to S. Thus in almost all practical applications the restriction will be valid and the modified estimates will be superior to x.

8

The restriction can be removed by a process outlined in the appendix provided k is sufficiently large ($k \geq 9$ is sufficient). This means that the proposed estimator will indeed dominate x for these values of k. The modifications to the estimator which are required are:

(i) r_k^* replaced by r_0^*.

(ii) S defined to be the greater of the values S, as originally defined, and $(p - N_0)/k$.

The proof above does little to illuminate the comparison of the two estimators. We suspect that as the number of non-zero observations increases, or as k increases, the superior choice is the modification above. However, this question has not been examined in this paper.

In earlier work k has always been assumed to be an integer. While the extension of the proofs of Section 2 for shrinkage towards non-integer, and unequal, hypothesized values seems fairly direct, the notational complexity required is forbidding, and we do not think it worthy of development. The adjustments required may best be illustrated by an example. Let $\kappa_1, \ldots, \kappa_p$, non-integer and unequal, be the hypothesized values of the means. To define the appropriate estimator corresponding to (2.3.1) we require values $\kappa_1^*, \ldots, \kappa_p^*$ in the transformed space, intuitively corresponding to $H(\kappa_1), \ldots, H(\kappa_p)$, undefined but easily obtained by extrapolation. (The actual values selected for $\kappa_1^*, \ldots, \kappa_p^*$ will not affect the proofs.) The estimate of the i^{th} co-ordinate of λ is then

$$\hat{\lambda}_i = x_i - \frac{r_\kappa^*}{S}(H(x_i) - \kappa_i^*), \text{ for } i = 1, \ldots, p,$$

where S is $\sum_{i=1}^{p} (H(x_i) - \kappa_i^*)^2$, and r_κ^* is the number of observations exceeding their hypothesized values, less three.

3. COMPUTER SIMULATION

In this section we report the results of a computer simulation used to compare our estimator $\hat{\lambda}$ given by (2.2.1) with the MLE. The simulation was performed on the Amdahl 470 V/6-II computer at the University of British Columbia. First, the number p of independent Poisson distributions is chosen. Second, p Poisson means are generated randomly within a certain range (c, d). Third, one observation of each of the p distributions with the means obtained in the second step is generated. Estimates of the means are then calculated using $\hat{\lambda}$. The third step is repeated 500 times and the risks for the estimator $\hat{\lambda}$ and the MLE are calculated. The percentage of savings in using $\hat{\lambda}$ as compared to the MLE $100[R(\lambda,X) - R(\lambda,\hat{\lambda})][R(\lambda,X)]^{-1}\%$, is calculated. The whole process is then repeated a number of times and the average percentage of the savings is calculated. We chose the range of the Poisson means in such a way that we might check the performance of the estimator $\hat{\lambda}$ when the means are relatively close to one another. This is the favorable situation for using $\hat{\lambda}$.

From Table 1, the improvement percentage is seen to be an increasing function of p, the number of independent Poisson distributions. Moreover, the improvement percentage increases as the magnitude of the λ_i's increases. From the computer simulation reported in Tsui's report, Peng's estimator given by (1.2) with $k = 0$ has a noticeable improvement over the MLE only when the Poisson means are in the interval (0,4). When the Poisson means are large, the improvement of his estimator is negligible. Our estimator $\hat{\lambda}^{(m)}$, however, is seen to have a considerable improvement over the MLE, especially when the λ_i's are large but close to one another.

Table 1. Improvement Percentage of $\hat{\lambda}^{(m)}$ over the MLE

Range of the Poisson Means	Percentage of Improvement over the MLE p = 5	p = 8	p = 10
(0,4)	4.2	7.9	10.4
(4,8)	6.1	13.0	14.0
(8,12)	7.3	14.2	15.4

4. AN APPLICATION

Efron & Morris (1975) cite simulation results for the estimation of binomial proportions which were expected to be near 0.05 (being the exact level at sample size N of a test with nominal significance level 0.05). Five hundred simulations were performed to estimate each proportion. The observed successes (rejections) for 17 different sample sizes are tabulated below as X. The distribution of X is (approximately) Poisson, with mean μ exactly calculable, and shown below.

j	X	μ	log X/25	δ_H
1	41	38.4	.495	36.4
2	21	25.1	-.174	22.6
3	23	21.1	-.083	23.8
4	20	26.4	-.223	22.1
5	27	32.0	.077	26.3
6	42	37.8	.519	37.2
7	18	20.5	-.329	21.1
8	18	22.8	-.329	21.1
9	20	25.8	-.223	22.1
10	25	28.8	.000	25.0
11	39	32.6	.445	34.9
12	15	26.5	-.511	19.8
13	18	21.3	-.329	21.1
14	30	22.9	.182	28.3
15	26	24.5	.039	25.6
16	23	27.1	-.083	23.8
17	27	29.8	.077	26.3

Calculations: (fixing $k = 25$) $S_H = 1.506$, $r_H^* = 14$.

$N = j+7$

The squared error of δ_H is 205 compared to the loss 449 of the naive estimator. The accuracy of δ_H is thus of the same order as the precision obtained by using the naive estimate based on over 1,000 simulations. (All but four co-ordinate estimates are improved, one of which is unchanged in value.)

We note, too, that the squared error of δ_H is much lower than that of the pooled mean (which is worse in this case than the naive estimator X). It is also smaller than the loss (242) obtained by Efron and Morris using Normal based Stein methods. This could be a chance fluctuation, or it could reflect the fact that the particular configuration of the μ's (being skew distributed rather than symmetrically spread here) is better suited to δ_H than to the Stein estimator.

5. CONCLUDING REMARKS

The estimators considered in Section 2 have been shown to have substantially smaller risks than that of the traditional estimator, or the Peng estimator, for situations in which the unknown means are likely to be similar in value but not close to zero. The risk reductions for the estimators considered in the simulations of Section 3 are as large as 15%.

Much greater gains can be anticipated when, in (2.3.1), the multiplying "constant" r_k^* can be replaced by r_0^*, without resulting in deterioration of risk relative to the usual estimator. This gain is apparent in the example of Section 4. Such a choice has been justified above, though it remains to be seen how the corresponding estimator will compare with the others.

The results obtained are important in underlining a useful

choice of metric, for estimation purposes, when considering the deviation of observations from preconceived values. The appropriate measure is seen to be $\sum_{i=1}^{p} H_{x_i}^2$ with H as given in (2.3.1). This observation makes it possible to extend these results to contingency table estimation by shrinkage to "smoothed" fitted values. This methodology will be the subject of a later paper.

Finally, we are once again led to conclude that prior expectations which are supported by the data can lead to quite dramatic improvements in estimator accuracy.

* * * * * * * * * * *

ACKNOWLEDGEMENT

The authors wish to thank Dr. J. V. Zidek for his encouragement.

This research was partially supported by the National Research Council of Canada and the Office of Naval Research, U.S.A. Part of the work was done while the authors were visiting at the University of British Columbia. Generous provision of facilities there are greatly appreciated.

BIBLIOGRAPHY

[1] Brown, C. C. and Muenz, L. R. (1976). "Reduced mean square error estimation in contingency tables," J. Amer. Stat. Assn., 71, 176-181.

[2] Brown, L. D. (1977). "A heuristic method for determining admissibility of estimators with applications," Annals Stat., (to appear).

[3] Carter, G. M. and Rolph, J. E. (1974). "Empirical Bayes methods applied to estimating fire alarm probabilities," J. Amer. Stat. Assn., 69, 880-885.

[4] Clevenson, M. L. and Zidek, J. V. (1975). "Simultaneous estimation of the mean of independent Poisson laws," J. Amer. Stat. Assn., 70, 698-705.

[5] Efron, B. and Morris, C. (1975). "Data analysis using Stein's estimator and its generalizations," J. Amer. Stat. Assn., 70, 311-319.

[6] Hudson, H. M. (1978). "A natural identity for exponential families with applications in multi-parameter estimation," Annals Stat., 6, 478-484.

[7] James, W. and Stein, C. (1961). "Estimation with Quadratic loss," Proc. Fourth Berk. Symp. Math. Stat. Prob., 1, 361-379.

[8] Leonard, T. (1976). "Some alternative approaches to multi-parameter estimation," Biometrika, 63, 69-75.

[9] Peng, J. (1979). "Simultaneous estimation of the parameters of independent Poisson distributions," Annals Stat., (to appear).

[10] Stein, C. (1973). "Estimation of the mean of a multivariate normal distribution," Tech. Report No. 48, Dept. of Statistics, Stanford University, Stanford, Calif.

[11] Tsui, K. (1978). "Simultaneous estimation of several Poisson parameters under squared error loss," Tech. Report No. 37, Dept. of Statistics, University of California at Riverside, Riverside, California.

APPENDIX

Proof of 2.4.1

As noted in Section 2.3, the argument used in Hudson (1978), Lemma (5.6), generalizes directly to this situation.

Proof of 2.4.2

We shall show that $A_j \geq \frac{1}{2}(1 - 3H_j^2/S)$ for $j = 1, 2, \ldots, k$ such that $N_j \geq 1$.

Write

$$A_j = \frac{1 - \frac{H_j^2}{S}(1 + \beta_j)}{1 + \frac{H_j^2}{S}(\beta_j^2 - 1)} ,$$

where $\beta_j = H_{j-1}/H_j = (1 - 1/jH_j)$, $j < k$.

The denominator is necessarily positive when $j \leq k$. For $j \neq m$, the minimum non-zero observation, $H_j(H_j + H_{j-1}) < S$, so, assuming that $\beta_j \leq 2$,

$$A_j = \frac{S - H_j(H_j + H_{j-1})}{S - \frac{1}{j}(H_j + H_{j-1})} \geq \frac{1 - \frac{H_j^2}{S}(1 + \beta_j)}{1 + 3\frac{H_j^2}{S}} \geq \frac{1 - \frac{3H_j^2}{S}}{1 + \frac{3H_j^2}{S}} .$$

Therefore, if $3H_j^2/S \leq 1$, (2.4.2) is immediate. If $3H_j^2/S > 1$, the result is still correct since the lefthand side is positive and the righthand side negative. The assumption $\beta_j \leq 2$ is correct for all j except $k-1 \leq j \leq k$ if $k \geq 4$. For these two remaining cases there is no difficulty in establishing that $A_j \geq \frac{1}{2}$ and (2.4.2) as a consequence, provided S is bounded away from 0 by the plus-rule

restriction $S \geq (p - N_0)/k$, when $k \geq 4$.

Finally, when $j = m$, the above argument applies unless $H_m(H_m + H_{m-1}) > S$. Then

$$0 \geq \frac{S - H_m(H_m + H_{m-1})}{S - \frac{1}{m}(H_m + H_{m-1})} \geq \frac{S - H_m^2(1 + \beta_m)}{S - \frac{1}{mH_m} S}$$

$$= \frac{1}{\beta_m}\left[1 - \frac{H_m^2}{S}(1 + \beta_m)\right] .$$

Since, for $k \geq 4$, $1 \leq \beta_m \leq 2$, we have

$$(2 - \beta_m) \geq (2 - \beta_m)\frac{H_m^2}{S} ,$$

which, on rearrangement, yields

$$1 - \frac{H_m^2}{S}(1 + \beta_m) \geq \frac{1}{2}\beta_m(1 - 3\frac{H_m^2}{S}) .$$

(2.4.2) again holds for $k \geq 4$.

Case by case enumeration reveals that (2.4.2) holds when $k = 1, 2$, or 3.

Risk improvement of the modified estimators of Section 2.4

Outline of a proof that $\Sigma N_j A_j \geq \frac{1}{2}(\Sigma N_j - 3)$.

1. As noted it suffices to consider the situation with $3S_+/S > \sum_{j=k+1}^{\ell} N_j$. If, in addition, $H_l(H_l + H_{l-1}) < S$ then we may replace (2.4.1) by the stronger inequality

$$\sum_{j=k+1}^{\ell} N_j A_j \geq \left(\sum_{j=k+1} N_j - 2S_+/S\right) , \qquad (1)$$

also a consequence of Hudson's paper.

This, together with (2.4.2), gives the result immediately (noting that $\sum_{j=k+1}^{\ell} N_j$ must be either 1 or 2). Thus we need only consider the case $H_1(H_1 + H_{1-1}) > S$.

2. Now treat the case $\sum_{j=k+1}^{\ell} N_j = 1$, $S_+/S > \frac{1}{2}$ (a consequence of $H_1(H_1 + H_{1-1}) > S$).

If $H_m(H_m + H_{m-1}) < S$ then show that for $k \geq 4$

$$\sum_{j=m}^{k-2} N_j A_j \geq \frac{M_1\left(1 - \frac{3R_1}{M_1}\right)}{1 + \frac{3R_1}{M_1}} \tag{2}$$

where $M_1 = \sum_{j=m}^{k-2} N_j A_j$, $R_1 = \sum_{j=m}^{k-2} N_j H_j^2/S$. This result follows by observing that, since $\beta_j \leq 2$, $A_j \geq \frac{1 - 3H_j^2/S}{1 + 3H_j^2/S}$, and that the minimum of the function

$$f(x) = \sum_{j=m}^{k-2} N_j \frac{1 - 3x_j}{1 + 3x_j}$$

for fixed $\underset{\sim}{N}$ and subject to the constraints $\Sigma N_j x_j = r$ occurs when $x_m = \ldots = x_{k-2} = r/M_1$.

Let $M_2 = \sum_{j=k-1}^{k} N_j$, $M_3 = \sum_{j=k+1}^{\ell} N_j (= 1)$ and $R_2 = \sum_{j=k-1}^{k} N_j H_j^2/S$, $R_3 = S_+/S$. Apply (2.4.1) and (2) to conclude

$$\Sigma N_j A_j \geq (1 - 3R_3) + M_2 A_{k-1} + \frac{M_1\left(1 - \frac{3R_1}{M_1}\right)}{1 + \frac{3R_1}{M_1}} .$$

Note that, for sufficiently large k, A_{k-1} is arbitrarily close to 1 because of the constraint that S exceed $(p-N_0)/k$. In particular, for $k \geq 9$, $A_{k-1} > 5/6$. Considering the righthand side as a function of R_1 and R_3 the minimum occurs when $1 + \frac{3R_1}{M_1} = \sqrt{2}$ and $R_3 = 1 - R_1$. Thus

$$\Sigma N_j A_j - \frac{(\Sigma N_j - 3)}{2} \geq (2\sqrt{2} - 2.5)M_1 + (A_{k-1} - 0.5)M_2 - 1$$

$$\geq 0 \quad \text{if} \quad A_{k-1} \geq 5/6 ,$$

since $M_1 + M_2 \geq 3$, except in the case $M_1 = 3$, $M_2 = 0$. A slight refinement of (2) shows the result for this case too.

If $H_m(H_m + H_{m-1}) > S$ then use the result

$$A_m \geq \frac{1}{2}\left(1 - \frac{3H_m^2}{S}\right)$$

from the proof of (2.4.2) together with the analogue of (2) for the remaining observations less than (k-1) to easily obtain the required result.

3. Now consider the case $\sum_{j=k+1}^{\ell} N_j = 2$, $S_+/S \geq 2/3$.

Since we may insist that $\beta_m < 2$, the only case that need be considered is $H_m(H_m + H_{m-1}) < S$. Thus (2) again applies, and repeating the argument above

$$\Sigma N_j A_j - \frac{(\Sigma N_j - 3)}{2} \geq (2\sqrt{2} - 2.5)M_1 + 0.33M_2 - 0.5 \geq 0 .$$

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Tech. Report #145, Series 2

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): SIMULTANEOUS POISSON ESTIMATORS FOR A PRIORI HYPOTHESES ABOUT MEANS

5. TYPE OF REPORT & PERIOD COVERED: Technical Report

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): H. Malcolm Hudson and Kam-Wah Tsui

8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0453

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, Princeton University, Princeton, N. J. 08544

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research (Code 436), Arlington, Virginia 22217

12. REPORT DATE: March 1979

13. NUMBER OF PAGES: 21

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES: Note: H. Malcolm Hudson is a Visiting Lecturer at Princeton University, on leave from the School of Economic and Financial Studies, Macquarie University, N.S.W., Australia. Kam-Wah Tsui is with the Dept. of Mathematics, Simon Fraser University, Vancouver, B.C., Canada

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

Estimators are presented for the simultaneous estimation of p means of Poisson variables with the properties of significant reduction in mean square error of estimation in the region of a pre-specified set of values and, if $p \geq 3$, safety, in the sense that mean square error is never worse than that of the minimum variance unbiased estimator.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)